

IMPLEMENTASI CASE BASE REASONING MENGGUNAKAN METODE COSINE SIMILARITY UNTUK MENDIAGNOSA PENYAKIT PADA SAPI

Sainah P. Faransyah¹, Sebastianus A. S. Mola, S. T., M.kom², Yelly Y. Nabuasa, S.Kom., M.Cs³

^{1,2,3}Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana

^{1,2,3}Email: saynaputri@gmail.com

INTISARI

Case Based Reasoning (CBR) merupakan teknik pemecahan kasus berdasarkan pengalaman pada kasus-kasus yang pernah terjadi sebelumnya dengan nilai kemiripan tertinggi. Pada penelitian ini, penulis menerapkan CBR untuk mendiagnosa penyakit pada sapi. Sumber pengetahuan sistem diperoleh dengan mengumpulkan data rekam medis pada tahun 2014, 2016, dan 2017. Sistem menggunakan metode *Rough Set* pada proses *indexing* dan untuk perhitungan nilai kemiripan menggunakan metode *Cosine Similarity* dengan nilai batas kewajaran adalah 70%. Sistem ini dapat mendiagnosa 15 penyakit berdasarkan 29 gejala yang ada. Keluaran sistem berupa diagnosa penyakit beserta solusi pengobatan dan presentasi kemiripan dengan kasus terdahulu untuk menunjukkan tingkat kebenaran hasil diagnosa. Berdasarkan hasil pengujian dengan menggunakan 30 kasus pada *casebase* didapatkan keakuratan sistem pada cara kedua adalah 27% dan pada cara ketiga didapat hasil terbaik dengan menggunakan 3 *fold* sebesar 33,33%. Sistem menghasilkan akurasi yang rendah karena jumlah kasus yang masih sedikit dan data yang tersebar pada kasus yang ada.

Kata Kunci: *Case-Based Reasoning, Rough Set, Cosine Similarity, Penyakit pada Sapi*

ABSTRACT

Case Based Reasoning (CBR) is a case-breaking technique based on experience in cases that have previously occurred with the highest similarity value. In this study, the authors apply CBR to diagnose cow disease. Sources of system knowledge are obtained by collecting cases from medical records on 2014, 2016, and 2017. The system uses the *Rough Set* method for indexing and the calculation of similarity values using the *Cosine Similarity* method with threshold 70%. This system is able to diagnose 15 diseases based on 29 existing symptoms. The output of the system in the form of the illness experienced, the solution and the presentation of similarities with the previous case to show the truth level of possible diagnose. Based on the test of 30 cases on *casebase* obtained system accuracy at second part is 27% and at third part the system gets the best result using 3 *fold* by 33,33%. The system produces low accuracy due to the small number of cases and the scattered data in the case.

Keyword: *Case Base Reasoning, Rough Set, Cosine Similarity, Cow Disease.*

I. PENDAHULUAN

Salah satu hewan ternak yang paling banyak ditanakkan adalah sapi. Sapi memberikan banyak manfaat dan keuntungan, antara lain dapat dimanfaatkan dagingnya, kulitnya, dan sebagainya. Sama dengan hewan ternak lain, sapi mempunyai bermacam jenis penyakit. Penyakit pada sapi merupakan suatu masalah kesehatan hewan di Indonesia khususnya di Provinsi Nusa Tenggara Timur (NTT). Penyakit tersebut disebabkan oleh banyak hal seperti virus, bakteri, parasit, cacing, dan sebagainya. Masalah penyakit dalam usaha peningkatan produksi ternak sapi merupakan gangguan dan ancaman yang serius. Kerugian yang ditimbulkan penyakit sapi dapat berbentuk kematian, pertumbuhan terlambat atau penurunan produksi.

Penyakit pada sapi tidaklah asing bagi peternak, namun merupakan kendala yang sering dihadapi peternak. Para peternak kadang tahu bila ternaknya terserang penyakit, tetapi peternak tidak tahu penyakit apa yang sedang menyerang ternaknya serta bagaimana mengatasinya. Sehingga untuk mengetahui penyakit yang sedang menyerang ternak dibutuhkan seorang pakar

yang dalam hal ini adalah dokter hewan. Namun pada kenyataannya para pakar atau dokter tidak selalu ada ketika dibutuhkan karena terbatasnya tenaga ahli. Dari hasil wawancara dengan drh. Theresia F.I.M.D Murni diketahui bahwa jumlah dokter hewan yang ada di Kabupaten Kupang yaitu sebanyak 11 dokter hewan. Sedangkan populasi ternak sapi sebanyak 186.553 [1]. Maka rasio dokter hewan terhadap populasi ternak sapi 11:186.553. Hal ini bila tidak segera ditangani besar kemungkinan jika ternak terserang penyakit maka ternak yang lain dapat tertular bahkan dapat berujung pada kematian.

Solusi untuk mengatasi uraian masalah diatas adalah dengan membuat suatu media yang menerapkan sistem kerja Case Based Reasoning berbasis web. Sistem Case Based Reasoning merupakan salah satu alternatif yang memudahkan peternak sapi dalam mendiagnosa penyakit sapi secara mandiri berdasarkan anjuran yang diberikan oleh sistem tersebut. CBR merupakan sebuah cara untuk menyelesaikan suatu kasus baru dengan cara mengadaptasi atau mengingat solusi yang terdapat pada kasus sebelumnya yang mirip dengan kasus baru tersebut [2].

Penelitian terdahulu menggunakan metode Cosine Similarity yaitu Sistem Deteksi Dini Diagnosa ISPA (Infeksi Saluran Pernapasan Akut) Pada Anak Dengan Metode Cosine Similarity [3], Pemanfaatan Metode Vector Space Model dan Metode Cosine Similarity pada Fitur Deteksi Hama dan Penyakit Tanaman Padi [4], dan Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia [5].

II. MATERI DAN METODE

2.1 Data Penelitian

Data yang dikumpulkan adalah data primer yang diambil dari Pusat Kesehatan Hewan (PUSKESWAN) Dinas Peternakan Kabupaten Kupang. Data yang diambil adalah data rekam medis pada penyakit sapi pada tahun 2014, 2016 dan 2017 dengan jumlah data sebanyak 115 kasus kemudian diolah dengan cara kasus-kasus tersebut dipilah sehingga diperoleh 30 data kasus yang tidak mengalami *redundant* atau kesamaan kasus dimana terdiri dari 15 penyakit dan 29 gejala.

2.2 Case Based Reasoning

Secara singkat, CBR merupakan salah satu teknik pemecahan masalah berdasarkan pengalaman-pengalaman pada masa lalu. Pengalaman yang pernah terjadi di masa lalu tersebut disimpan dalam tempat yang disebut dengan basis kasus. Aamodt dan Plaza [6], menjelaskan sebuah CBR sebagai sebuah siklus yang disingkat 4 R yaitu

1. *Retrieve* dilakukan untuk mendapatkan/memperoleh kembali kasus yang paling menyerupai/relevan (*similar*) dengan kasus yang baru.
2. *Reuse* dilakukan untuk memodelkan/menggunakan kembali pengetahuan dan informasi kasus lama berdasarkan bobot kemiripan yang paling relevan ke dalam kasus yang baru, sehingga menghasilkan usulan solusi dimana mungkin diperlukan suatu adaptasi dengan masalah yang baru tersebut.
3. *Revise* dilakukan untuk meninjau kembali solusi yang diusulkan kemudian mengetesnya pada kasus nyata (simulasi) dan jika diperlukan memperbaiki solusi tersebut agar cocok dengan kasus yang baru.
4. *Retain* dilakukan untuk mengintegrasikan/menyimpan kasus baru yang telah berhasil mendapatkan solusi agar dapat digunakan oleh kasus-kasus selanjutnya yang mirip dengan kasus tersebut.

2.3 Indexing

Indexing merupakan proses pengelompokkan kasus yang ada berdasarkan fitur yang ditentukan. Proses *indexing* dapat mengefisienkan waktu dan memory karena pada saat melakukan proses pencarian kasus yang mirip dengan kasus yang baru, sistem CBR tidak perlu menghitung nilai kemiripan kasus baru terhadap semua kasus yang ada, namun cukup menghitung

nilai kemiripan terhadap kasus yang berada pada kelompok yang sama dengan kasus baru tersebut. Pada penelitian ini metode yang digunakan untuk *indexing* adalah *rough set*.

2.4 Rough Set

Teori *rough set* adalah sebuah teknik matematik yang dikembangkan oleh Pawlack pada tahun 1980 [7]. *Rough Set* salah satu teknik data mining yang digunakan untuk menangani masalah *Uncertainty*, *Imprecision* dan *Vagueness* dalam aplikasi *Artificial Intelligence* (AI). *Rough Set* merupakan teknik yang efisien untuk *Knowledge Discovery in Database* (KDD) dalam tahapan proses dan Data Mining.

2.5 Cosine Similarity

Cosine similarity adalah perhitungan kesamaan antara dua vektor n dimensi dengan mencari kosinus dari sudut diantara keduanya dan sering digunakan untuk membandingkan dokumen dalam *text mining* [8]. *Cosine Similarity* memiliki kelebihan yaitu tidak terpengaruh pada panjang pendeknya suatu dokumen [9].

Rumus *Cosine similarity* adalah sebagai berikut:

$$\begin{aligned} \text{Similarity}(x,y) = \cos(\Theta) &= \frac{x \cdot y}{||x|| ||y||} \\ &= \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \cdot \sqrt{\sum_{i=1}^n (y_i)^2}} \dots\dots\dots(2.1) \end{aligned}$$

Dimana :

$x \cdot y$ = *vector dot product* dari x dan y , dihitung dengan $\sum_{i=1}^n x_i \cdot y_i$

$||x||$ = panjang vektor x dihitung dengan $\sum_{i=1}^n (x_i)^2$

$||y||$ = panjang vektor y dihitung dengan $\sum_{i=1}^n (y_i)^2$

$x_i = (x_1, x_2, x_3, \dots, x_n)$ = gejala – gejala pada kasus baru

$y_i = (y_1, y_2, y_3, \dots, y_n)$ = gejala – gejala pada kasus lama

n = jumlah gejala

2.6 Tahap Perancangan Sistem CBR

Perancangan sistem merupakan proses atau tahap untuk membangun sistem. Penelitian ini bertujuan untuk menerapkan CBR dengan menggunakan metode *Cosine Similarity* dalam mendiagnosa penyakit pada sapi. Gambaran tahapan CBR yaitu: *retrieve*, *reuse*, *revise* dan *retain* dalam sistem CBR diagnosa penyakit pada sapi secara garis besar dapat dilihat pada *flowchart* gambar 3.1.

III. HASIL DAN PEMBAHASAN

3.1 Hasil Pengujian

Sistem CBR telah diberikan pengujian dengan 3 (tiga) cara pengujian. Cara pertama, sistem diuji dengan memasukan gejala yang sama dengan gejala pada satu kasus dalam basis kasus. Cara kedua, sistem diuji dengan memasukan 30 kasus yang diujikan dengan 3 cara pengujian mendapatkan nilai akurasi pada pengujian pertama sebesar 20%, pada pengujian kedua sebesar 20% dan pada pengujian ketiga sebesar 40%. Rata-rata akurasi sebesar 27%. Cara ketiga, sistem diuji dengan memasukan 30 kasus yang diujikan dengan menggunakan *cross fold validation* didapatkan hasil rata-rata akurasi terbesar pada perhitungan dengan menggunakan 3 *fold* dengan nilai 33,33%.



3.2 Pembahasan

Perbandingan tingkat keakuratan sistem antara pengujian cara kedua dan pengujian cara ketiga menggunakan *cross fold validation* adalah berbeda. Akurasi yang rendah dihasilkan oleh sistem dikarenakan jumlah *case* yang masih sedikit dan penyebaran data pada kasus-kasus yang ada. Dari berbagai pembahasan di atas, dapat disimpulkan bahwa sistem dapat mendiagnosa penyakit pada sapi dengan tingkat keakuratan pengujian cara kedua dalam mendiagnosa penyakit pada sapi adalah sebesar 27% dan pengujian cara ketiga dengan 10 *fold* yaitu 27%, dengan 5 *fold* yaitu 23,33%, dan dengan 3 *fold* yaitu sebesar 33,33%.

IV. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian pada sistem *case base reasoning* untuk mendiagnosa penyakit pada sapi menggunakan metode *Cosine Similarity* berbasis web, dapat disimpulkan sebagai berikut :

1. Dari hasil pengujian tahap pertama diperoleh kesimpulan yaitu sistem berhasil mendiagnosa penyakit pada sapi serta memberikan solusi dengan nilai kemiripan tertinggi 100% dan jika ada kasus yang memiliki nilai kemiripan dibawah 70% maka akan mendapatkan revisi oleh pakar.
2. Pengujian pada cara kedua terhadap 30 kasus yang diujikan dengan 3 cara pengujian mendapatkan nilai akurasi pada pengujian pertama sebesar 20%, pada pengujian kedua sebesar 20% dan pada pengujian ketiga sebesar 40%. Rata-rata akurasi sebesar 27%.
3. Pengujian pada cara ketiga terhadap 30 kasus yang diujikan dengan menggunakan *cross fold validation* didapatkan hasil rata-rata akurasi terbesar pada perhitungan dengan menggunakan 3 fold dengan nilai 33,33%.
4. Penyakit yang memiliki nilai similaritas tinggi tetapi hasil diagnosa berbeda antara pakar dan sistem dikarenakan penyakit yang di uji tidak ada didata latih sehingga sistem mendiagnosa penyakit yang memiliki gejala yang paling mirip. Maka dari itu hasil perhitungan diagnosa sistem tidak sama dengan diagnosa pakar.
5. Akurasi yang rendah dihasilkan oleh sistem dikarenakan jumlah *case* yang masih sedikit dan penyebaran data pada kasus-kasus yang ada.
6. Secara umum sistem *Case Base Reasoning* dapat digunakan untuk melakukan pendiagnosaan penyakit pada hewan sapi.

4.2 Saran

Sebagai saran dari hasil penelitian ini adalah diharapkan peneliti selanjutnya dapat menambahkan jumlah basis kasus karena keterbatasan data sehingga kasus yang dibahas dalam penelitian hanya sebanyak 30 data kasus dengan 15 penyakit dan 29 data gejala saja serta untuk meningkatkan keakuratan sistem. Selanjutnya peneliti juga dapat menyempurnakan proses *indexing* dengan menggunakan metode *rough set* dalam pencarian *index* kasus.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik (2015), www.bps.go.id diakses pada 5 Juli 2017
- [2] Pal, S., 2004. *Foundations of Soft Case-Based Reasoning*. Hoboken: Wiley Interscience publication.
- [3] Cahyani E. D., 2013, Sistem Deteksi Dini Diagnosa ISPA (Infeksi Saluran Pernapasan Akut) Pada Anak Dengan Metode *Cosine Similarity*.
- [4] Triana A., Saptono R., Sulistyio E. M., -, Pemanfaatan Metode *Vector Space Model* dan Metode *Cosine Similarity* pada Fitur Deteksi Hama dan Penyakit Tanaman Padi.
- [5] Nurdiana O., Jumadi, Nursantika D., 2016, Perbandingan Metode *Cosine Similarity* dengan Metode *Jaccard Similarity* Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia, JOIN, Volume I No. 1.
- [6] Aamodt, A., Plaza, E. 1994. *Case Based Reasoning: Foundation Issues Methodological Variations, and System Approaches*, *AI Communication* Vol 7 Nr, 1 March 1994, pp 39-59.
- [7] Chouchoulas, A., 1999, "A Rough Set Approach To Text Classification", The University Of Edinburgh.
- [8] Zhiqiang, L., Werimin, S., Zhenhua, Y. (2009). *Measuring Semantic Similarity between Words Using Wikipedia*. IEEE. 251-255

- [9] Rozas, I. R. dan Sarno, R., 2011, SiPKoKI ISO 27001: Sistem Pemilihan Kontrol Keamanan Informasi Berbasis ISO 27001, *Seminar Nasional Pascasarjana XI*, ITS, Surabaya, 27 Juli